



# Semantic diversity in paired-associate learning: Further evidence for the information accumulation perspective of cognitive aging

Mengyang Qiu<sup>1</sup> · Brendan T. Johns<sup>1</sup>

Published online: 10 December 2019  
© The Psychonomic Society, Inc. 2019

## Abstract

Normal aging is often associated with a performance decline on various cognitive tests, including paired associate learning (PAL), where participants are asked to learn and recall arbitrary word pairs. While many studies have taken this as evidence to support the notion of age-related deficits in cognitive processing, Ramscar, Hendrix, Shaoul, Milin, and Baayen (*Topics in Cognitive Science*, 6(1), 5–42) and Ramscar, Sun, Hendrix, and Baayen (*Psychological Science*, 28(8), 1171–1179, 2017) posit that the decline in performance on various cognitive tasks can be explained by the accumulation of linguistic knowledge over time. To demonstrate this, Ramscar et al. (2017) found that older bilingual participants outperformed monolingual counterparts on a verbal PAL task, proposed to be due to bilinguals having accumulated less information about the words used in the study. However, comparing bilinguals to monolinguals introduces confounding factors. For example, bilingual's better performance may be due to superior executive functioning. To minimize these between-subject confounds, the current study used a within-subject design in order to examine the influence of linguistic experience on paired associate learning in younger and older adults. Linguistic experience was modeled using a semantic diversity measure of word strength (Jones, Johns, & Recchia, 2012). When frequency is controlled for, high semantic diversity words are associated to a greater number of words and have a higher average strength of association. In the current study, PAL performance of older adults was significantly lower for word pairs involving high semantic diversity words, while their performance did not differ for low semantic diversity words, consistent with the information accumulation perspective of aging.

**Keywords** Paired-associate learning · Semantic diversity · Memory · Corpus-based modeling · Information accumulation · Aging

Paired-associate learning (PAL) has long been used to assess memory performance across aging in both experimental and clinical settings (e.g., Naveh-Benjamin, 2000; Wechsler, 1945, 2009; Zaretsky & Halberstam, 1968). In a verbal PAL task, subjects are asked to memorize a list of arbitrary cue-response word pairs (e.g., *sky-tea*, *day-box*). At test they are asked to recall response words when provided with the cue words. The associative ability tested in PAL tasks is considered an essential mechanism for human memory (e.g., associating a name with a person; desRosiers & Ivison, 1986).

Many studies have shown that older adults consistently perform worse on PAL tasks, including nonverbal stimuli (e.g., *product-price*, *item-location*, *face-name*), compared

with younger adults (for a review, see Naveh-Benjamin & Mayr, 2018). Naveh-Benjamin (2000) attributed the worse performance in older adults to a deficit in their ability to bind new pieces of information together. Alternative theories have claimed that older adults' insufficient ability in inhibitory control (Hasher & Zacks, 1988) leads to binding too much irrelevant information (hyperbinding), and consequently interferes with forming relevant associations during learning and selecting target associations during retrieval in PAL tasks (e.g., Biss, Campbell, & Hasher, 2012; Campbell, Hasher, & Thomas, 2010). These theories have been used as evidence to support the notion that memory, along with other cognitive systems (e.g., attention and executive control), decline as one ages (Park & Festini, 2017).

Recent studies by Ramscar and colleagues (Blanco et al., 2016; Ramscar, Hendrix, Love, & Baayen, 2013; Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014; Ramscar, Sun, Hendrix, & Baayen 2017), however, have challenged this deficit perspective of cognitive aging. Ramscar and colleagues propose

✉ Brendan T. Johns  
btjohns@buffalo.edu

<sup>1</sup> Department of Communicative Disorders and Sciences, University at Buffalo, 122 Cary Hall, Buffalo, NY 14214, USA

that the poorer PAL performance seen in older adults reflects the accumulation of linguistic knowledge over time rather than any age-related cognitive decline. As a person ages, they necessarily experience more occurrences of meaningful word sequences, which increases their knowledge of the relationship between words. The accumulation of these meaningful associations will, in turn, cause unrelated words to become more negatively associated (Ramscar et al., 2017). Thus, forming associations between unrelated word pairs in PAL will be particularly problematic for older adults, simply because that “the learning of a nonsensical link between two unconnected words must increasingly compete with prior learning to the effect that this link is nonsensical” (Ramscar, Hendrix, Love, & Baayen, 2013, p. 458).<sup>1</sup> We will refer to the theories put forth by Ramscar and colleagues as the information accumulation perspective of aging (see also Wulff, De Deyne, Jones, Mata, & Aging Lexicon Consortium, 2019, for a similar perspective on aging).

Based on this notion, Ramscar et al. (2017) predicted that age-matched individuals who had less experience with language should perform better on a PAL task. To operationalize language exposure, Ramscar et al. (2017) used age-matched bilinguals and monolinguals. Bilingual speakers necessarily have less linguistic experience in one language compared with monolingual speakers of that language because their usage and experience with a language is split (Gollan et al., 2011). Thus, the information accumulation perspective of aging predicts that bilinguals should have less interference from prior linguistic knowledge. The results of Ramscar et al. (2017) bore this prediction out, with older bilinguals performing significantly better on a verbal PAL task than older monolingual participants.

However, comparing monolinguals to bilinguals introduces one potential confounding variable—namely, executive control. Previous studies have found that older bilinguals perform better than age-matched monolinguals on executive control, including better distraction inhibition (e.g., Bialystok, Craik, & Luk, 2008; Bialystok, Craik, & Ryan, 2006; Treccani, Argyri, Sorace, & Della Sala, 2009; for a review, see Bialystok, 2017). Although many recent studies have challenged the bilingual advantages in executive functioning (e.g., Dick et al., 2019; Lehtonen et al., 2018; Von Bastian, Souza, & Gade, 2016), most of these studies focus on young adults, the population who typically have peak performance in various executive functioning tasks. The effects have been observed more consistently in older adults (for a review, see Antoniou, 2019). Thus, the better performance of bilinguals found in Ramscar et al. (2017) may also be taken as evidence to support the inhibitory deficit hypothesis of paired-associate learning (e.g., Campbell et al., 2010). That is, the better

executive functioning of bilinguals helps them inhibit irrelevant associations, which older monolinguals may find difficult, and thus promotes PAL performance.

The goal of the current study is to remove this potential confound by using a within-subjects design, where it will be shown that older adults can perform equally as well as younger adults when the number and strength of associations that a word has is controlled for. To accomplish this, the semantic diversity model (SDM; Jones et al., 2012) was used to quantify the associative strength that a word has to other words in the lexicon, described below.

## Semantic diversity as a measure of association

The semantic diversity (SD) of a word measures the content variability of the contexts that the word occurs in. The greater the number of dissimilar contexts that a word appears in, the higher the word’s semantic diversity. Jones et al. (2012) showed that a semantic diversity measure provided a superior fit to lexical decision and naming times over a word frequency and contextual diversity count (operationalized as the number of documents or paragraphs that a word occurs in; see Adelman, Brown, & Quesada, 2006).

Subsequent studies have further supported the importance of semantic diversity across various tasks, including spoken word recognition (Johns, Gruenenfelder, Pisoni, & Jones, 2012), natural language learning (Johns, Dye, & Jones, 2016a), and word recognition across aging and bilingualism (Johns, Sheppard, Jones, & Taler, 2016b). All these studies have demonstrated that the semantic diversity of the contexts that words appear in are an important organizational principle of our mental lexicon. Words that appear in more unique contexts are more likely to be needed in new contexts, and thus, are stored more strongly in memory (for a review, see Jones, Dye, & Johns, 2017).

Additionally, contextual and semantic diversity has been shown to be an important information source in child language acquisition and usage (e.g., Hills, Maouene, Riordan, & Smith, 2010; Hsiao & Nation, 2018; Hsiao, Bird, Norris, Pagán, & Nation, 2019; Joseph & Nation, 2018).

Although semantic diversity was not originally built to measure associations between words, it captures this information indirectly. A word that is of high semantic diversity necessarily appears in more distinct contexts across learning, which means that it also co-occurs with a greater number of words. Oppositely, a low semantic diversity word appears in redundant contexts, and thus only co-occurs with a more limited set of words. Given equal frequency of two words, a high semantic diversity word will have a proportionally greater number of associations with other words compared with a word with low semantic diversity. The below simulation using the SDM will validate these assumptions.

<sup>1</sup> The same learning principle is also supported by results from other nonverbal PAL tasks—for example, older adults have difficulty associating items with unrealistic prices (competing with prior knowledge), but not realistic ones (Castel, 2005; Fine, Shing, & Naveh-Benjamin, 2018). The current study focused on verbal PAL because prior knowledge of word associations can be more precisely quantified using corpus-based models.

The SDM is a computational model that measures the semantic diversity of words within a corpus by weighting the uniqueness of information that each new context provides, using an expectancy-congruence mechanism (Jones et al., 2012). To accomplish this, words are represented in a Word  $\times$  Context matrix. A context is operationalized as groups of 20 sentences from a corpus of natural language. Each time the model encounters a new context, a new column is added to the matrix. For words that do not occur in the new context, its value for that column is zero. For each word that occurs in that context, its value for that column is determined by taking the similarity between the current context and the stored meaning of that word (the word's corresponding row in the matrix). The more dissimilar the current context is compared with past usage, the higher the value is in that column. A word's semantic diversity value is then the sum of the values in its row (for a formal description of the model, see Jones et al., 2012).

To formally establish the relationship between semantic diversity and word association, we ran the SDM on a set of words with equal frequency but different semantic diversity, and examined how frequency, semantic diversity, number of associations, and strength of associations change as a function of the number of contexts the model experiences. To measure these variables, the SDM was run on 100,000 contexts, from a corpus of young adult books (see Johns, Jones, & Mewhort, 2019). Number of associations was defined as the number of unique words that the target words co-occurred with across learning (equivalent to the degree of a node in a semantic network). Average association strength was calculated by going through each word that a target word co-occurred with and summing the SD values across their joint contexts, and then dividing it by the total number of words that the target word co-occurred with. This will give the average association strength that a word has to other words in the lexicon. The target words used in the simulation are described in the Materials section below, attained from the young adult corpus (see Appendix Table 2).

Figure 1 displays the results of this simulation. Figure 1a shows that the two sets of words are matched on word frequency, while Fig. 1b shows that the word sets diverge on semantic diversity. Figure 1c shows that words with high semantic diversity both occur with a greater number of words (i.e., these words have more associations), and, importantly, Fig. 1d shows that they also have a higher average level of association to those words. This means that words that are high in semantic diversity have both a greater number of associations to other words in the lexicon, and also a stronger level of association to those words.

From a learning theory perspective, this suggests that high SD words should be relatively more difficult to form new associations to. However, the current simulation also shows that the divergences that SD causes takes greater amounts of experience: as the model accumulates experience, there is a greater separation between high and low SD words. This suggests that

older adults should be more affected by high SD words, given the greater level of experience that older adults have had with language. The goal of this article is to test this prediction by giving older and younger adults a paired-associate learning task, where the cue word is either low or high in semantic diversity. If the information accumulation perspective on aging is correct, then older adults should perform significantly worse when needing to form new associations to high SD words compared with low SD words, while younger adults should perform relatively equally to low and high SD words.

## Method

### Participants

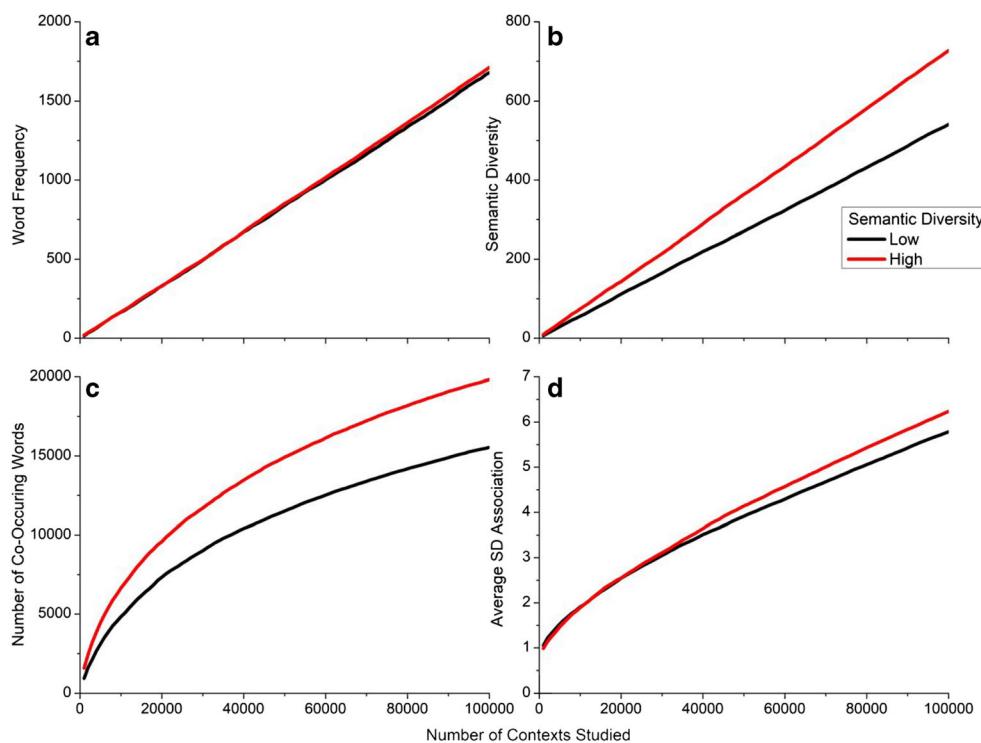
Participants were younger (18–29 years old) and older (45–60 years old) native speakers of American English recruited from Prolific.ac, an online subject pool for behavioral studies (Palan & Schitter, 2018). Each participant was awarded \$2.71 for their participation. A sample size of at least 50 participants in each age group was decided prior to data collection based on previous PAL studies. Fifty-seven younger and 57 older participants completed the experiment. We excluded data from five younger participants and seven older participants because either their overall accuracy was less than 25% or they got 0% accuracy in one of the experimental trials. Thus, data from 52 younger (age  $M = 23$  years,  $SD = 3$  years) and 50 older (age  $M = 52$  years,  $SD = 5$  years) participants were analyzed.

### Materials

Cue and response words were selected from two corpora—one composed of young adult books and the other composed of general fiction books (see Johns & Jamieson, 2019; Johns et al., 2019). Each corpus consisted of 100,000 contexts, operationalized with a moving window of 20 sentences (meaning each corpus consisted of 2,000,000 sentences). The general fiction corpus contained approximately 29.2 million words, while the young adult fiction contained approximately 25.6 million words. The size discrepancy between the two corpora is due to young adult books having shorter sentences.

Ten high semantic diversity and 10 low semantic diversity words with equal frequency were selected as cue words from each corpus. Twenty response words of similar frequency were also selected from each corpus, with semantic diversity measures in between the high and low SD cue words. All the words were also matched on familiarity and imageability from the MRC Psycholinguistic Database (Coltheart, 1981). The detailed statistics of the stimuli is shown in Table 1, and the specific words used are contained in the Appendix Table 2.

The experiment is a 2 (older vs. younger adults)  $\times$  2 (high vs. low SD) factorial experiment. For every subject, each cue



**Fig. 1** Various measurements of word strengths and word association from the semantic diversity model. **a** Shows that low and high semantic diversity word sets have equal word frequency. **b** Shows that they diverge on semantic diversity. **c** Shows that high semantic diversity words have a

greater number of overall associations. **d** Shows that they also have a greater average strength of association. Values are means from the high and low semantic diversity words contained in the Appendix Table 2 for the young adult corpus

word was randomly paired with a different response word. The resulting 40 word pairs were randomly assigned to four experimental lists of 10 pairs, counterbalanced across subjects. Each list contained five high and five low semantic diversity word pairs. Words from different corpora do not intermix in the same list (i.e., two lists contained the word pairs from the young adult corpus and two lists contained the word pairs from the general fiction corpus).

## Procedure

The experiment was implemented using jsPsych (de Leeuw, 2015), and data collection was managed by JATOS (Lange,

**Table 1** Statistics of the stimuli (reported as mean  $\pm$  standard deviation)

	Freq (log)	SD (log)	Familiarity	Imageability
Young adult fiction				
High SD cue	8.61 $\pm$ 0.12	7.68 $\pm$ 0.09	544 $\pm$ 30	562 $\pm$ 41
Low SD cue	8.60 $\pm$ 0.14	7.28 $\pm$ 0.14	541 $\pm$ 45	583 $\pm$ 33
Response	8.56 $\pm$ 0.10	7.51 $\pm$ 0.03	552 $\pm$ 36	577 $\pm$ 35
General fiction				
High SD cue	8.62 $\pm$ 0.14	7.64 $\pm$ 0.08	557 $\pm$ 27	568 $\pm$ 52
Low SD cue	8.59 $\pm$ 0.13	7.42 $\pm$ 0.03	537 $\pm$ 57	570 $\pm$ 39
Response	8.48 $\pm$ 0.11	7.50 $\pm$ 0.02	535 $\pm$ 38	534 $\pm$ 63

Note. SD = semantic diversity

Kühn, & Filevich, 2015). There were four experimental trials, each containing a learning phase and a test phase of a word-pair list. During the learning phase, each word pair in a list appeared twice in the center of the screen in random order. Each word pair stayed on screen for 1.5 seconds. A cross was displayed in between word pairs for 1 second. Participants were instructed to memorize all the word pairs. Immediately after the learning phase, there was a cued recall test, in which participants were provided with one cue word at a time in randomized order, and they were asked to enter the other word in that pair. Before the experimental trials, each participant also finished a practice trial with five word pairs selected from the original Wechsler Memory Scale (Wechsler, 1945) in order for them to be familiar with the task.

## Results

There was no performance difference for the word pairs from the young adult or general fiction corpus, so the performance of two word lists were collapsed and analyzed together.

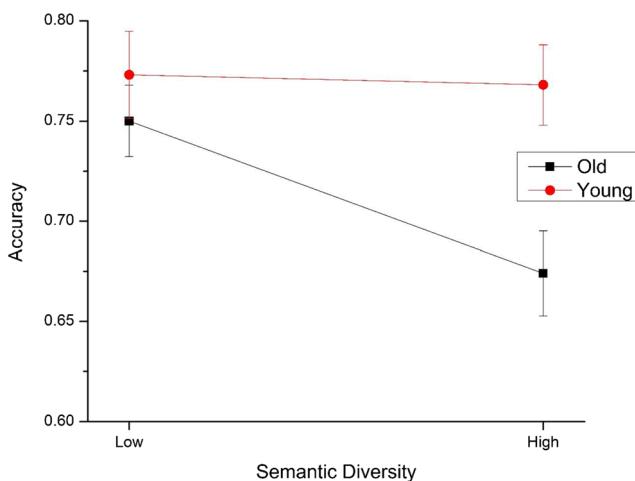
We conducted generalized linear mixed-effects models on the effects of age (younger vs. older), level of semantic diversity (low vs. high) on the accuracy of paired-associate learning using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2018). Age and level of semantic diversity were

coded with contrast coding (1, -1), and were included as fixed effects, both as main effects and as interactions. We chose the model with the maximal random effect structure that would converge justified by the data (Barr, Levy, Scheepers, & Tily, 2013). Specifically, participants were included as a random intercept and a random slope by semantic diversity uncorrelated with the random intercept. Items were included as a random intercept only. The complete model was  $GLMM\_ACC = glmer(Accuracy \sim Age \times SD + (1|Participants) + (0 + SD|Participants) + (1|Item), family = "binomial," data = PAL\_data)$ . Significance was assessed via model comparison with an alpha of 0.05.

The data are displayed in Fig. 2. A significant effect of age was found, such that the accuracy of older participants was lower than that of younger participants,  $\beta = 0.22$ ,  $SE = 0.10$ ,  $\chi^2(1) = 4.84$ ,  $p < .05$ . Importantly, there was a significant interaction of age and semantic diversity,  $\beta = -0.11$ ,  $SE = 0.04$ ,  $\chi^2(1) = 8.03$ ,  $p < .01$ . For older adults, there was a significant effect of semantic diversity,  $\beta = 0.21$ ,  $SE = 0.06$ ,  $\chi^2(1) = 11.32$ ,  $p < .001$ , meaning that their accuracy was lower in the high SD condition than in the low SD condition. However, the effect of SD was not significant within younger participants,  $\beta = -0.01$ ,  $SE = 0.07$ ,  $\chi^2(1) = 0.02$ ,  $p = .88$ . More importantly, when comparing the two age groups, we found that there was no significant performance difference in the low semantic diversity condition,  $\beta = 0.09$ ,  $SE = 0.10$ ,  $\chi^2(1) = 0.86$ ,  $p = .35$ . The significant effect of age was driven by the worse performance of older adults in the high semantic diversity condition,  $\beta = 0.33$ ,  $SE = 0.11$ ,  $\chi^2(1) = 8.79$ ,  $p < .01$ .

## Discussion

The goal of this article was to further evaluate the information accumulation perspective of aging using paired-



**Fig. 2** Results of the verbal PAL task for older and younger adults with cue words that were either high or low in semantic diversity. The data indicate that the older adults have a greater difficulty in forming new associations to high semantic diversity words. Error bars are standard error

associate learning, building off of the work of Ramscar et al. (2017). This perspective proposes that it is the continual accumulation of experience in memory that leads older adults to have worse performance across a variety of psychometric tests (Ramscar et al. 2014). Ramscar et al. (2017) explicitly tested this hypothesis using age-matched bilinguals and monolinguals. The goal of the current study was to remove the potential confound of executive functioning differences between monolinguals and bilinguals by using a within-subjects design. This was accomplished by attaining words that have either high or low levels of association to other words, but equivalent word frequency, through the use of semantic diversity (Jones et al., 2012). In a simulation study, it was shown that words high in semantic diversity had a greater level of association to other words, compared with words that had low semantic diversity (see Fig. 1). It was found that even though older adults had lower overall accuracy, the decrease in performance was driven by the high SD words. The performance of older adults did not differ from younger adults in the low SD condition. This finding demonstrates that it is the greater amount of associations that older adults have acquired through experience that causes their poorer performance on a PAL task.

The results of this paper show the power and promise of using corpus-based models to better understand trends in development and aging, joining a growing body of literature (e.g., Hills et al., 2010; Hsiao & Nation, 2018; Johns et al., 2018; Taler, Johns, & Jones, 2019). Corpus-based models allow for a quantification of the lexical information that people have been exposed to at various stages of aging. By mapping model outputs to human behavior, it provides insights into how experience shapes cognition. When combined with large-scale data collection designed to estimate the quantity and type of lexical experience (e.g., Brysbaert, Stevens, Mandera, & Keuleers, 2016), these models provide a promising pathway for the development of more realistic cognitive models of aging.

The results contained in this article, as well as the work of Ramscar et al. (2014, 2017), suggest that the behavioral detriments that are seen in older adults are not necessarily due to any structural changes in the cognitive system of older adults, but instead reflect differences in informational task difficulty that older adults face. In terms of a PAL task, the buildup of word associations within the lexicon impairs the ability of older adults to form arbitrary new associations. Corpus-based modeling allows for a determination of the buildup of lexical information across the life span, enabling a better understanding of the impact of information accumulation on older adults.

**Open practices statement** The data for this experiment are available at <https://osf.io/4rdzg/>.

## Appendix

**Table 2** Note: This data is mandatory. Please provide.

General fiction	Condition	Log-WF	Log-SD	Imageability	Familiarity
Library	Low-SD	8.52	7.45	587	580
Lawyer	Low-SD	8.62	7.45	557	520
Engine	Low-SD	8.47	7.45	595	543
Commander	Low-SD	8.69	7.44	478	409
Lake	Low-SD	8.55	7.44	616	583
Blade	Low-SD	8.51	7.43	568	517
Hood	Low-SD	8.86	7.42	558	510
University	Low-SD	8.50	7.39	615	622
Aunt	Low-SD	8.71	7.38	567	554
Secretary	Low-SD	8.44	7.37	563	528
Cigarette	High-SD	8.66	7.61	645	573
Trip	High-SD	8.52	7.61	520	559
Cheek	High-SD	8.47	7.61	561	533
Sweat	High-SD	8.61	7.67	560	545
Color	High-SD	8.84	7.80	513	582
Jacket	High-SD	8.55	7.60	611	596
Entrance	High-SD	8.60	7.62	493	555
Hospital	High-SD	8.86	7.76	602	548
Mountain	High-SD	8.57	7.58	629	574
Horror	High-SD	8.47	7.57	545	501
Chain	Response	8.53	7.54	559	513
Wound	Response	8.46	7.54	570	474
Deck	Response	8.58	7.47	539	507
Clock	Response	8.57	7.53	614	608
Card	Response	8.61	7.52	578	543
Wave	Response	8.48	7.51	542	518
Whisper	Response	8.33	7.51	567	550
Bird	Response	8.57	7.51	614	592
Dirt	Response	8.44	7.50	547	571
Van	Response	8.80	7.50	572	542
Roll	Response	8.37	7.50	496	547
Crime	Response	8.53	7.50	471	537
Dance	Response	8.49	7.49	510	550
Silk	Response	8.40	7.49	510	482
Prison	Response	8.47	7.48	593	462
Style	Response	8.34	7.48	416	555
Motion	Response	8.46	7.50	416	541
Appearance	Response	8.39	7.47	425	578
Grave	Response	8.39	7.47	619	501
Club	Response	8.46	7.45	522	533
Young Adult Fiction	Condition	Log-WF	Log-SD	Imageability	Familiarity
Cabin	Low-SD	8.69	7.40	582	523
Professor	Low-SD	8.53	7.03	587	583
Spear	Low-SD	8.50	7.35	545	513
Wolf	Low-SD	8.87	7.37	610	537
Milk	Low-SD	8.44	7.20	638	588
Chief	Low-SD	8.45	7.14	545	482

**Table 2** (continued)

Truck	Low-SD	8.47	7.12	621	620
Ash	Low-SD	8.74	7.35	553	494
Valley	Low-SD	8.65	7.45	600	515
Plane	Low-SD	8.63	7.36	556	558
Chin	High-SD	8.87	7.91	608	545
Pleasure	High-SD	8.41	7.57	511	583
Nature	High-SD	8.61	7.71	513	535
Shell	High-SD	8.66	7.71	581	524
Pile	High-SD	8.64	7.68	513	521
Spring	High-SD	8.55	7.63	585	588
Band	High-SD	8.56	7.63	579	555
Dawn	High-SD	8.57	7.63	586	507
Laughter	High-SD	8.66	7.71	613	570
Flash	High-SD	8.59	7.65	528	509
Crystal	Response	8.65	7.55	579	510
Cell	Response	8.52	7.53	590	520
Team	Response	8.50	7.48	565	538
Country	Response	8.55	7.51	539	592
Station	Response	8.63	7.53	554	548
Plate	Response	8.46	7.51	527	556
Map	Response	8.70	7.49	587	545
Park	Response	8.60	7.48	573	571
Letter	Response	8.66	7.46	595	610
Cloth	Response	8.47	7.54	547	561
Beach	Response	8.66	7.54	667	553
Meal	Response	8.41	7.47	573	603
Bar	Response	8.51	7.49	596	592
Beard	Response	8.46	7.55	630	538
Soldier	Response	8.54	7.52	578	517
Arrow	Response	8.73	7.49	619	490
Mud	Response	8.55	7.52	582	519
Princess	Response	8.73	7.52	547	502
Kid	Response	8.47	7.52	525	559
Coat	Response	8.48	7.54	572	610

## References

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, 17(9), 814–823.

Antoniou, M. (2019). The advantages of bilingualism debate. *Annual Review of Linguistics*, 5, 395–415.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). doi:<https://doi.org/10.18637/jss.v067.i01>

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, 143(3), 233–262. doi:<https://doi.org/10.1037/bul000009>

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 859–873.

Bialystok, E., Craik, F. I., & Ryan, J. (2006). Executive control in a modified antisaccade task: Effects of aging and bilingualism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(6), 1341–1354. doi:<https://doi.org/10.1037/0278-7393.32.6.1341>

Biss, R. K., Campbell, K. L., & Hasher, L. (2012). Interference from previous distraction disrupts older adults' memory. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 68(4), 558–561.

Blanco, N. J., Love, B. C., Ramscar, M., Otto, A. R., Smayda, K., & Maddox, W. T. (2016). Exploratory decision-making as a function of lifelong experience, not cognitive decline. *Journal of Experimental Psychology: General*, 145(3), 284–297. doi:<https://doi.org/10.1037/xge0000133>

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, 7, 1116. doi:<https://doi.org/10.3389/fpsyg.2016.01116>

Campbell, K. L., Hasher, L., & Thomas, R. C. (2010). Hyper-binding: A unique age effect. *Psychological Science*, 21(3), 399–405.

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging*, 20(4), 718.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 33A(4), 497–505.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12.

desRosiers, G., & Ivison, D. (1986). Paired associate learning: Normative data for differences between high and low associate word pairs. *Journal of Clinical and Experimental Neuropsychology*, 8 (6), 637–642.

Dick, A. S., Garcia, N. L., Pruden, S. M., Thompson, W. K., Hawes, S. W., Sutherland, M. T., ... Gonzalez, R. (2019). No evidence for a bilingual executive function advantage in the nationally representative ABCD study. *Nature Human Behavior*, 1, 692–701. doi:<https://doi.org/10.1038/s41562-019-0609-3>

Fine, H. C., Shing, Y. L., & Naveh-Benjamin, M. (2018). Effects of changes in schematic support and of item repetition on age-related associative memory deficits: Theoretically-driven empirical attempts to reduce older adults' high false alarm rate. *Psychology and Aging*, 33(1), 57–73. doi:<https://doi.org/10.1037/pag0000211>

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, 140(2), 186–209.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 22, pp. 193–225). New York, NY: Elsevier.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, 63(3), 259–273.

Hsiao, Y., Bird, M., Norris, H., Pagán, A., & Nation, K. (2019). The influence of item-level contextual history on lexical and semantic judgments by children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language*, 103, 114–126.

Johns, B. T., Dye, M., & Jones, M. N. (2016a). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, 23(4), 1214–1220.

Johns, B. T., Gruenenfelder, T. M., Pisoni, D. B., & Jones, M. N. (2012). Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *The Journal of the Acoustical Society of America*, 132(2), 74–80.

Johns, B. T., & Jamieson, R. K. (2019). The influence of time and place on lexical behavior: A distributional analysis. *Behavior Research Methods*. Advance online publication. doi:<https://doi.org/10.3758/s13428-019-01289-z>

Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2019). Using experiential optimization to build lexical representations. *Psychonomic Bulletin & Review*, 26(1), 103–126.

Johns, B. T., Sheppard, C. L., Jones, M. N., & Taler, V. (2016b). The role of semantic diversity in word recognition across aging and bilingualism. *Frontiers in Psychology*, 7, 1–11.

Johns, B. T., Taler, V., Pisoni, D. B., Farlow, M. R., Hake, A. M., Kareken, D. A., ... Jones, M. N. (2018). Cognitive modeling as an interface between brain and behavior: Measuring the semantic decline in mild cognitive impairment. *Canadian Journal of Experimental Psychology*, 72(2), 117–126.

Jones, M. N., Dye, M., & Johns, B. T. (2017). Context as an organizing principle of the lexicon. In B. J. Ross (Ed.), *Psychology of learning and motivation* (Vol. 67, pp. 239–283). New York, NY: Academic Press.

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, 66(2), 115–124.

Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology*, 166, 190–211.

Lange, K., Kühn, S., & Filevich, E. (2015). "Just another tool for online studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, 10(6), 1–14.

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, 144(4), 394–425. doi:<https://doi.org/10.1037/bul0000142>

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning Memory and Cognition*, 26(5), 1170–1187.

Naveh-Benjamin, M., & Mayr, U. (2018). Age-related differences in associative memory: Empirical evidence and theoretical perspectives. *Psychology and Aging*, 33(1), 1–6.

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27.

Park, D. C., & Festini, S. B. (2017). Theories of memory and aging: A look at the past and a glimpse of the future. *The Journals of Gerontology: Series B*, 72(1), 82–90.

R Core Team. (2018). R: A language and environment for statistical computing [Computer software]. Retrieved from <https://www.r-project.org>

Ramscar, M., Hendrix, P., Love, B., & Baayen, R. (2013). Learning is not decline: The mental lexicon as a window into cognition across the lifespan. *The Mental Lexicon*, 8(3), 450–481.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, 6(1), 5–42.

Ramscar, M., Sun, C. C., Hendrix, P., & Baayen, H. (2017). The mismeasurement of mind: Life-span changes in paired-associate-learning scores reflect the "cost" of learning, not cognitive decline. *Psychological Science*, 28(8), 1171–1179.

Taler, V., Johns, B. T., & Jones, M. N. (2019). A large scale semantic analysis of verbal fluency across the aging spectrum: Data from the Canadian longitudinal study on aging. *Journal of Gerontology: Psychological Sciences*. Advance online publication. doi:<https://doi.org/10.1093/geronb/gbz003>

Treccani, B., Argyri, E., Sorace, A., & Della Sala, S. (2009). Spatial negative priming in bilingualism. *Psychonomic Bulletin & Review*, 16(2), 320–327.

Von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology: General*, 145(2), 246–258. doi:<https://doi.org/10.1037/xge0000120>

Wechsler, D. (1945). *Wechsler Memory Scale*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2009). *Wechsler Memory Scale–Fourth Edition (WMS-IV)*. San Antonio, TX: Pearson.

Wulff, D. U., De Deyne, S., Jones, M. N., Mata, R., & Aging Lexicon Consortium. (2019). New perspectives on the aging lexicon. *Trends in Cognitive Sciences*, 23(8), 686–698. doi:<https://doi.org/10.1016/j.tics.2019.05.003>

Zaretsky, H. H., & Halberstam, J. L. (1968). Age differences in paired-associate learning. *Journal of Gerontology*, 23(2), 165–168.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.